УДК 004.8

А. Л. Колпаков
Научный руководитель: канд. тех. наук, доц. А. Н. Соколов
Южно-Уральский государственный университет, Челябинск

ПРИМЕНЕНИЕ АЛГОРИТМОВ ИНТЕЛЛЕКТУАЛЬНОГО АНАЛИЗА ДАННЫХ В СИСТЕМАХ ОБНАРУЖЕНИЯ ВТОРЖЕНИЙ

Аннотация. Рассмотрены алгоритмы интеллектуального анализа данных: метод опорных векторов, метод *k*-ближайших соседей, дерево принятия решений, нейронная сеть. Приведен сравнительный анализ преимуществ и недостатков рассмотренных алгоритмов, сделаны выводы по проблематике их использования для построения систем обнаружения вторжений.

Ключевые слова: информационная безопасность; защита информации; метод опорных векторов; метод *k*-ближайших соседей; дерево принятия решений; нейронная сеть.

В основе действия системы обнаружения атак лежит модель, с помощью которой предполагается сетевой трафик относить либо к нормальной сетевой активности, либо к аномальной.

Было проведено сравнение методов интеллектуального анализа данных.

1. Метод опорных векторов (Support Vector Machine, SVM) — набор схожих алгоритмов обучения с учителем, применяющихся в задачах классификации и регрессионного анализа.

Основная идея метода — перевод исходных векторов в пространство более высокой размерности и поиск разделяющей гиперплоскости с максимальным зазором в этом пространстве. Две параллельных гиперплоскости строятся по обеим сторонам гиперплоскости, разделяющей классы. Разделяющей гиперплоскостью будет гиперплоскость, максимизирующая расстояние до двух параллельных гиперплоскостей. Алгоритм работает в предположении, что чем больше разница или расстояние между этими параллельными гиперплоскостями, тем меньше будет средняя ошибка классификатора [1].

В качестве достоинств SVM можно отметить способность к обобщению, высокую точность и низкую вычислительную сложность принятия решения. Недостатком метода является относительно большая вычислительная сложность построения классифицирующей модели [2].

2. Метод *k*-ближайших соседей (*k*-nearest neighbors, *k*-NN) — метрический алгоритм для автоматической классификации объектов или регрессии.

Классифицируемый объект относится к тому классу, которому принадлежат ближайшие к нему объекты обучающей выборки [3].

Метод *k*-NN является одним из наиболее простых методов ИАД. Результаты применения метода легко поддаются интерпретации. Недостаток метода — его чувствительность к локальной структуре данных [2].

3. Дерево принятия решений — средство поддержки принятия решений, использующееся в статистике и анализе данных для прогнозных моделей [4].

Структура дерева представляет собой «листья» и «ветки». На ребрах («ветках») дерева решения записаны атрибуты, от которых зависит целевая функция, в «листьях» записаны значения целевой функции, а в остальных узлах — атрибуты, по которым различаются случаи. Чтобы классифицировать новый случай, надо спуститься по дереву до листа и выдать соответствующее значение. Подобные деревья решений широко используются в интеллектуальном анализе данных. Цель состоит в том, чтобы создать модель, которая предсказывает значение целевой переменной на основе нескольких переменных на входе.

Достоинствами деревьев принятия решений являются простой принцип их построения и хорошая интерпретируемость результатов, недостатком — невысокая точность классификации [2].

4. Нейронная сеть (или искусственная нейронная сеть) — математическая модель, а также ее программное или аппаратное воплощение, построенная по принципу организации и функционирования биологических нейронных сетей — сетей нервных клеток живого организма. Нейронные сети не программируются в привычном смысле этого слова, они обучаются. Обучение сети происходит путем корректировки значений весов нейронов для минимизации ошибки классификации [5].

Таблица 1

Сравнение алгоритмов интеллектуального анализа данных

Алгоритм	Точность	Масштабируемость	Трудоемкость	Быстрота	Популярность
Метод опорных векторов	Высокая	Высокая	Высокая	Высокая	Высокая
Метод *k*-ближайших соседей	Низкая	Очень низкая	Нейтральная/ Низкая	Очень низкая	Низкая
Дерево принятия решений	Низкая	Высокая	Высокая	Высокая/ Нейтральная	Высокая/ Нейтральная
Нейронные сети	Высокая	Низкая	Низкая	Низкая	Низкая

Преимущества нейронных сетей выражаются в их способности автоматически приобретать знания в ходе обучения, а также способности к обобщению, основной недостаток состоит в чувствительности к шуму во входных данных [2].

В итоге все алгоритмы интеллектуального анализа данных можно сравнить между собой, оценивая характеристики их свойств.

Как видно из данных табл. 1, каждый алгоритм имеет свои сильные и слабые стороны, но ни один метод не способен решить весь спектр задач интеллектуального анализа данных.

Список литературы

1. *Nefedov A.* Support Vector Machines: A Simple Tutorial // svmtutorial.online. URL: https://svmtutorial.online/.

2. *Шарабыров И. В.* Система обнаружения атак в локальных беспроводных вычислительных сетях на основе технологий интеллектуального анализа данных : дис. … канд. тех. наук. Уфа, 2016. 144 с.

3. *Hastie T.*, *Tibshirani R.*, *Friedman J.* The Elements of Statistical Learning. Springer, 2001.

4. Microsoft Decision Trees Algorithm // Электронный портал «Microsoft». URL: https://docs.microsoft.com/en-us/sql/analysis-services/data-mining/microsoft-decision-trees-algorithm.

5. *Уоссермен Ф.* Нейрокомпьютерная техника: теория и практика. М. : Мир, 1992.

УДК 004.056.53

Д. А. Корепин
Научный руководитель: д-р тех. наук, проф. С. В. Поршнев
Уральский федеральный университет, Екатеринбург

ИССЛЕДОВАНИЕ ВОЗМОЖНОСТЕЙ ВОССТАНОВЛЕНИЯ ИНФОРМАЦИИ В ФАЙЛОВОЙ СИСТЕМЕ REFS

Аннотация. В настоящей статье изучены и проанализированы официальные материалы по описанию файловой системы ReFS и механизмов ее работы. Рассмотрены возможности восстановления информации в данной файловой системе.

Ключевые слова: восстановление данных; файловая система; ReFS.

Файловая система ReFS была представлена в 2012 году [1]. С момента выпуска данной файловой системы прошло 5 лет, однако подробной документации